

523,332

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2004年2月5日 (05.02.2004)

PCT

(10) 国際公開番号
WO 2004/012101 A1

(51) 国際特許分類: G06F 17/30

(21) 国際出願番号: PCT/JP2003/009678

(22) 国際出願日: 2003年7月30日 (30.07.2003)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願2002-221698 2002年7月30日 (30.07.2002) JP

(71) 出願人 (米国を除く全ての指定国について): ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒141-0001 東京都品川区北品川6丁目7番35号 Tokyo (JP).

(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 木村 仁史 (KIMURA, Hitoshi) [JP/JP]; 〒141-0001 東京都品川区北品川6丁目7番35号 ソニー株式会社内 Tokyo (JP). 大沼 顕介 (OHNUMA, Kensuke) [JP/JP];

〒141-0001 東京都品川区北品川6丁目7番35号 ソニー株式会社内 Tokyo (JP). 市岡 秀俊 (ICHIOKA, Hidetoshi) [JP/JP]; 〒141-0001 東京都品川区北品川6丁目7番35号 ソニー株式会社内 Tokyo (JP).

(74) 代理人: 角田 芳末, 外 (TSUNODA, Yoshisue et al.); 〒160-0023 東京都新宿区西新宿1丁目8番1号 新宿ビル Tokyo (JP).

(81) 指定国 (国内): CN, KR, US.

(84) 指定国 (広域): ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR).

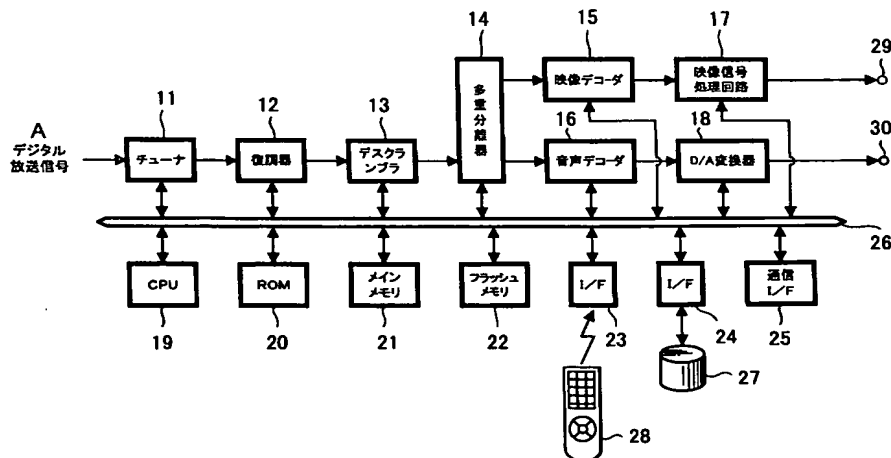
添付公開書類:

— 国際調査報告書

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(54) Title: AUTOMATIC KEYWORD EXTRACTION DEVICE AND METHOD, RECORDING MEDIUM, AND PROGRAM

(54) 発明の名称: キーワードの自動抽出装置及び方法、記録媒体、並びにプログラム



A... DIGITAL BROADCAST SIGNAL

11... TUNER

12... DEMODULATOR

13... DE-SCRAMBLER

14... MULTIPLEXER/DEMULTIPLEXER

15... VIDEO DECODER

16... AUDIO DECODER

17... VIDEO SIGNAL PROCESSING CIRCUIT

18... D/A CONVERTER

21... MAIN MEMORY

22... FLASH MEMORY

25... COMMUNICATION I/F

(57) Abstract: An automatic keyword extraction device includes first extraction means (19) for extracting a keyword from content title character string information by using a first keyword dictionary containing character strings indicating sub-genres and second extraction means (19) for extracting a keyword from content detailed character string information by using a second keyword dictionary containing person names and extracting a keyword by utilizing the *zisyukiri* method (font delimiting method).

/続葉有/-

WO 2004/012101 A1



Thus, in a home electric appliance having a small CPU processing ability and memory capacity, it is possible to effectively, accurately, and automatically extract a keyword for a user to search a content from content title character string information and detailed character string information such as EPG information.

(57) 要約: コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第1のキーワード辞書を用いてキーワードの抽出を行う第1の抽出手段19と、このコンテンツの詳細文字列情報から、人名が登録されている第2のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第2の抽出手段19とを備える。これにより、CPUの処理能力やメモリの容量がさほど大きくない家電製品でも、EPG情報のようなコンテンツのタイトル文字列情報及び詳細文字列情報から、ユーザーがコンテンツを検索するためのキーワードを、効率よく且つ精度よく自動的に抽出できるようにする。

明 細 書

キーワードの自動抽出装置及び方法、記録媒体、並びにプログラム

技術分野

- 5 本発明は、E P G (Electronic Program Guide : 電子番組ガイド) 情報のようなコンテンツのタイトル文字列情報及び詳細文字列情報から、キーワードを自動的に抽出する装置及び方法、記録媒体、並びにプログラムに関する。

10 背景技術

- 近年本格化しているデジタルテレビジョン放送では、番組の映像・音声データとともに、番組のタイトルを示す情報（タイトル文字列情報）や番組の詳細を説明する情報（詳細文字列情報）や番組のジャンルを示す情報等を含んだE P G情報が、放送局から送信される。デジタル放送に対応したテレビジョン受信機では、このE P G情報に基づいて画面上に電子番組ガイドを表示させることができる。

また、アナログテレビジョン放送でも、こうしたE P G情報が送信されているものがある。

- 20 ユーザーは、見たい番組を検索する場合、この電子番組ガイドを利用して、大まかなジャンル（例えばスポーツ、ドラマ等）を選んだ後、タイトルから検索したり、詳細文字列情報を読んで検索したりしている。

- 25 しかし、番組のタイトルの付け方は千差万別であるので、タイトルからの検索はユーザーにとって必ずしも行いやすいものではない。また、番組の詳細文字列情報は文章の形式で記述されており何ページにも亘ることが少なくないので、詳細文字列情報からの検索もユーザーにとって面倒である。

これに対し、例えば芸能人名等のキーワードを用いて番組を検索できるようにすれば、ユーザーにとって検索が非常に容易になる。しかるに、現在放送局から送信されるE P G情報には、キーワードは独立して含まれていない。そのため、キーワードを用いた検索を可能にするためには、E P G情報からキーワードを抽出することが必要になる。

従来、このキーワードの抽出方法としては、テレビジョン受信機に表示された電子番組ガイド中の詳細文字列情報の文章の中から、ユーザーが、キーワードとして決定したい文字列の先頭及び
10 末尾の語をカーソル等で指定するという方法が存在していた。

しかし、この従来の抽出方法では、ユーザーが自らキーワード指定のための操作を行わなければならないので、煩雑であるとともに、多数のキーワードを短時間に抽出することは困難である。

他方、一般的なキーワードの自動抽出方法としては、日本語形態要素解析という方法が知られている。しかし、この方法は、プログラムのサイズや使用する辞書のサイズが非常に大きいとともに、C P Uに大きな負荷をかけてしまう。したがって、テレビジョン受信機のようにC P Uの処理能力やメモリの容量がさほど大きくない家電製品でこの方法を用いることは、極めて非効率的である。
15 20

さらに、一般的なキーワードの自動抽出方法としては、字種切り法という方法も知られている。この方法は、漢字・カタカナ・平仮名・アルファベット・数字等の字種の違いを検出することによってキーワードを抽出するものである。しかし、この字種切り
25 法だけでは、番組を検索するためのキーワードの抽出を精度よく行うことはできない。すなわち、苗字が漢字で名前が平仮名やカタカナの芸能人名（例えば‘石田あかり’というような名称）は、苗字と名前とが分割されてしまうので抽出することができない。

また、名前がアルファベットで表記され苗字がカタカナで表記された外国人名や名前と苗字との間に‘・’（中点）が挿入された外国人名（例えば‘Ｂ・ドゥーリー’というような名称）も、名前と苗字とが分割されてしまうので抽出することができない。

- 5 本発明は、上述の点に鑑み、ＣＰＵの処理能力やメモリの容量がさほど大きくない家電製品でも、ＥＰＧ情報のようなコンテンツのタイトル文字列情報及び詳細文字列情報から、ユーザーがコンテンツを検索するためのキーワードを、効率よく且つ精度よく自動的に抽出できるようにすることを課題としてなされたものである。
- 10

発明の開示

- この課題を解決するために、本出願人は、コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている
- 15 第１のキーワード辞書を用いてキーワードの抽出を行う第１の抽出手段と、このコンテンツの詳細文字列情報から、人名が登録されている第２のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第２の抽出手段とを備えたキーワード自動抽出装置を提案する。

- 20 このキーワード自動抽出装置では、コンテンツのタイトル文字列情報（例えばテレビジョン放送ではＥＰＧ情報中のタイトル文字列情報）からは、サブジャンルを示す文字列が登録されている第１のキーワード辞書を用いてキーワードが抽出される。

- また、コンテンツの詳細文字列情報（例えばテレビジョン放送
- 25 ではＥＰＧ情報中の詳細文字列情報）からは、人名が登録されている第２のキーワード辞書を用いてキーワードが抽出されるとともに、字種切り法を利用したキーワードの抽出も行われる。その際、第２のキーワード辞書に登録されている人名であれば、苗字

が漢字で名前が平仮名やカタカナの人名もキーワードとして抽出される。また、第2のキーワード辞書に登録されていない人名でも、字種切り法を利用することによってキーワードとして抽出される。

- 5 このように、タイトル文字列情報からのキーワードの抽出と詳細文字列情報からのキーワードの抽出とを、それぞれの情報に合わせて互いに異なるキーワード辞書とルール（字種切り法を利用するか否か等）で行うことにより、小さなサイズのプログラムや辞書で精度よくキーワードを抽出することができる。
- 10 これにより、CPUの処理能力やメモリの容量がさほど大きくない家電製品でも、EPG情報のようなコンテンツのタイトル文字列情報及び詳細文字列情報から、ユーザーがコンテンツを検索するためのキーワードを、効率よく且つ精度よく自動的に抽出することができるようになる。
- 15 なお、このキーワード自動抽出装置において、一例として、第1の抽出手段は、第1のキーワード辞書に登録されている文字列を含むタイトル文字列のうち、所定の除外文字列辞書に登録されている文字列を除外した部分からキーワードを抽出することが好適である。
- 20 それにより、タイトルに含まれていることのある文字列のうち、コンテンツを検索するためには不適切な（一般的過ぎるような）文字列がキーワードに含まれることを防止することができる。したがって、ユーザーは、抽出されたキーワードを用いて、コンテンツを一層効率的に検索することができるようになる。
- 25 さらに、このキーワード自動抽出装置において、一例として、第1の抽出手段は、第1のキーワード辞書に登録されている文字列を含むタイトル文字列のうち、平仮名、カタカナ、漢字、数字、アルファベット以外の特殊文字で区切られている文字列をキーワ

ードとして抽出することが好適である。

それにより、こうした特殊文字で区切られていないタイトルについては、そのタイトルに含まれる複数の文字列がばらばらのキーワードとして抽出されることなく、そのタイトルそのものがその
5 のままの形でキーワードとして抽出されるようになる。

こうした特殊文字で区切られていないタイトルは、そのタイトルに含まれる個々の文字列は意味が広すぎてコンテンツ検索のためのキーワードとしてあまり役立たず（検索結果が多くなり）、タイトルそのものとしてはじめてコンテンツの効率的な検索
10 のためのキーワードとして役立つことが多い。したがって、ユーザーは、抽出されたキーワード（タイトルそのもの）を用いて、コンテンツを一層効率的に検索することができるようになる。

また一方では、特殊文字で区切られているタイトルについては、特殊文字で区切られている個々の文字列がそれぞれキーワードとして抽出されるようになる。
15

特殊文字（例えばスペースや‘×’等）で区切られているタイトルは、その特殊文字で区切られている個々の文字列がそれぞれコンテンツ検索のためのキーワードとして役立ち、タイトルそのものとしては限定されすぎてコンテンツ検索のためのキーワード
20 としてあまり役立たない（検索結果がゼロまたは非常に少なくなる）ことが多い。したがって、ユーザーは、抽出されたキーワード（特殊文字で区切られている個々の文字列）を用いて、やはりコンテンツを一層効率的に検索することができるようになる。

さらに、このキーワード自動抽出装置において、一例として、
25 第2の抽出手段は、第2のキーワード辞書を用いてキーワードを抽出した詳細文字列情報の残りの部分のうち、所定の除外文字列辞書に登録されている文字列を除外した部分から、字種切り法を利用したキーワードの抽出を行うことが好適である。

それにより、詳細文字列情報に含まれていることのある文字列のうち、コンテンツを検索するためには不適切な文字列がキーワードに含まれることを防止することができる。したがって、ユーザーは、抽出されたキーワードを用いて、コンテンツを一層効率的に検索することができるようになる。

さらに、このキーワード自動抽出装置において、一例として、第2の抽出手段は、字種切り法を利用しつつ、カタカナとアルファベットとを同一字種として扱うとともに、‘・’（中点）は、その直前の文字がカタカナ、アルファベットである場合にはそれぞれカタカナ、アルファベットとして扱うことが好適である。

それにより、名前がアルファベットで表記され苗字がカタカナで表記された外国人名や名前と苗字との間に‘・’（中点）が挿入された外国人名も、キーワードとして抽出することができるようになる。

さらに、このキーワード自動抽出装置において、第2のキーワード辞書をネットワーク経由でダウンロードする手段をさらに備え、第2の抽出手段はこのダウンロードされた第2のキーワード辞書を用いることが好適である。

それにより、第2のキーワード辞書として、最新の辞書（最近有名になったばかりの人の名称も登録されている辞書）を用いてキーワードを抽出することができるようになる。

次に、本出願人は、コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第1のキーワード辞書を用いてキーワードの抽出を行う第1のステップと、このコンテンツの詳細文字列情報から、人名が登録されている第2のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第2のステップとを有するキーワード自動抽出方法を提案する。

また、キーワード自動抽出装置のプログラムであって、コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第１のキーワード辞書を用いてキーワードの抽出を行う第１の抽出ステップと、このコンテンツの詳細文字列情報から、人名が登録されている第２のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第２の抽出ステップとを含むコンピュータが読み取り可能なプログラムが記録されている記録媒体を提案する。

また、キーワード自動抽出装置を制御するコンピュータに、コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第１のキーワード辞書を用いてキーワードの抽出を行う第１の抽出ステップと、このコンテンツの詳細文字列情報から、人名が登録されている第２のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第２の抽出ステップとを実行させるプログラムを提案する。

このキーワード自動抽出方法や記録媒体やプログラムによれば、前述の本発明に係るキーワード自動抽出装置について説明したのと全く同様にして、ＣＰＵの処理能力やメモリの容量がさほど大きくない家電製品でも、ＥＰＧ情報のようなコンテンツのタイトル文字列情報及び詳細文字列情報から、ユーザーがコンテンツを検索するためのキーワードを、効率よく且つ精度よく自動的に抽出することができるようになる。

図面の簡単な説明

図１は、本発明を適用した番組記録再生装置を含むデジタルテレビジョン放送受信システムの概要を示す図である。

図２は、図１の番組記録再生装置のハードウェア構成を示すブロック図である。

図 3 は、図 2 の C P U が実行するキーワードの自動抽出処理を示すフローチャートである。

図 4 は、図 2 の C P U が実行するキーワードの自動抽出処理を示すフローチャートである。

5 図 5 は、図 3 の処理におけるキーワード抽出のためのルールを示す図である。

図 6 は、図 4 の処理におけるキーワード抽出のためのルールを示す図である。

10 図 7 は、本発明を適用したアナログテレビジョン放送用の番組記録再生装置のハードウェア構成を示すブロック図である。

発明を実施するための最良の形態

以下、デジタルテレビジョン放送の番組を記録・再生する装置に本発明を適用した例について、図面を用いて説明する。

15 図 1 は、本発明を適用した番組記録再生装置を含むデジタルテレビジョン放送受信システムの概要を示す図である。テレビジョン放送局から送信されたデジタル放送信号が、アンテナ 1 で受信されて番組記録再生装置 2 に入力する。番組記録再生装置 2 は、ディスプレイ及びスピーカを含む表示装置 3 に接続されるととも
20 に、インターネット 4 に接続されている。

図 2 は、番組記録再生装置 2 のハードウェア構成を示すブロック図である。この番組記録再生装置 2 では、チューナ 1 1，復調器 1 2，デスクランブラ 1 3，多重分離器 1 4 が順に接続されるとともに、多重分離器 1 4 に対して映像デコーダ 1 5，映像信号
25 処理回路 1 7 と音声デコーダ 1 6，D/A 変換器 1 8 とがそれぞれ順に接続されている。

また、チューナ 1 1～D/A 変換器 1 8，C P U 1 9，R O M 2 0，メインメモリ（R A M）2 1，フラッシュメモリ 2 2，リ

モートコントローラ用のインターフェース 23, HDD (ハードディスクドライブ) 用のインターフェース 24, インターネット接続用の通信インターフェース 25 が、互いにシステムバス 26 で結ばれている。インターフェース 24 には、テレビ番組を録画するための HDD (ハードディスクドライブ) 27 が接続されている。

この番組記録再生装置 2 に付属したリモートコントローラ (以下リモコンと呼ぶ) 28 には、通常のデジタル放送用テレビジョン受信機に付属したリモコンにおけるのと同じ各種の操作釦 (電源釦や、選局釦や、録画予約釦や、再生釦や、EPG画面上で選択を行うための方向キーや決定キー等) が設けられている。

テレビ番組の視聴時には、番組記録再生装置 2 に入力したデジタル放送信号は、リモコン 28 の選局操作に基づいてチューナ 11 で周波数帯を選択された後、復調器 12 で復調され、デスクランブラ 13 でスクランブルを解かれた後、多重分離器 14 で、複数チャンネル分の番組の映像・音声データの packets や EPG 情報の packets に分離される。

この複数チャンネル分のテレビ番組の映像・音声の packets のうち、リモコン 28 の選局操作に基づいて抽出した 1 チャンネル分の packets の映像、音声データが、それぞれ映像デコーダ 15, 音声デコーダ 16 により MPEG-2 Video, MPEG-2 Audio で復号される。また、EPG 情報の packets は CPU 19 に送られる。

そして、映像デコーダ 15 で復号された映像信号や、EPG 情報を用いて CPU 19 で作成された電子番組ガイド表示用の映像信号が、映像信号処理回路 17 で NTSC 方式への変換やミキシング等を施され、映像出力端子 29 から出力して図 1 の表示装置 3 に送られる。

また、音声デコーダ 16 で復号された音声信号が、D/A 変換器 18 でアナログ変換され、音声出力端子 30 から出力して図 1 の表示装置 3 に送られる。

5 CPU 19 は、ROM 20 に格納されたプログラムやデータに基づき、メインメモリ 21 をワーキングメモリとして用いてこの番組記録再生装置 2 全体を制御する。

10 CPU 19 が行う処理には、リモコン 28 の選局操作に基づくテレビ番組の視聴時の処理や、リモコン 28 の録画予約操作に基づく HDD 27 へのテレビ番組の録画処理の他に、キーワードの自動抽出処理がある。

ROM 20 には、このキーワードの自動抽出処理で用いるための辞書として、タイトル用キーワード辞書、タイトル用除外文字列辞書、詳細情報用キーワード辞書及び詳細情報用除外文字列辞書が格納されている。

15 タイトル用キーワード辞書には、‘プロ野球’、‘ゴルフ’、‘サッカー’、‘温泉’、‘囲碁’、‘将棋’、‘映画’等のサブジャンル（EPG 情報中のジャンル情報による‘スポーツ’といったような大まかなジャンルよりも細かいジャンル）を示す文字列や、‘恋’、‘愛’といった文字列や、プロ野球の球団名の文字列のように、
20 番組のタイトルに含まれていることの多い文字列のうち、番組を検索するために有効且つ重要な文字列が登録されている。

タイトル用除外文字列辞書には、‘映画’、‘BS’、番組表独特の記号（例えば、ニュース番組を表す、Nを四角の枠で囲んだ記号）といったように、番組のタイトルに含まれていることのある
25 文字列のうち、番組を検索するためのキーワードとしては一般的過ぎる文字列が登録されている。

詳細情報用キーワード辞書には、テレビジョン番組に登場することの多い有名人（芸能人、スポーツ選手、政治家、文化人等）

の名称のうち、平仮名のみ、平仮名と漢字との組み合わせ、平仮名とカタカナとの組み合わせ、漢字とカタカナとの組み合わせ、2文字以下の漢字のみ、6文字以上の漢字のみの名称の文字列がそれぞれ登録されている。また、詳細情報用キーワード辞書には、

5 例えば‘温泉’というような、E P G情報中の詳細文字列情報に含まれていることの多い文字列のうち、人名以外の文字列であって番組を検索するためのキーワードとして適切な文字列も登録されている。

詳細情報用除外文字列辞書には、‘ゲスト’，‘以上’，‘監督’の

10 ように、E P G情報中の詳細文字列情報に含まれていることの多い文字列のうち、番組を検索するためのキーワードとしては不適切な文字列が登録されている。

なお、C P U 1 9は、詳細情報用キーワード辞書に関しては、インターネット経由で専用のサイトから最新のもの（最近有名になったばかりの人の名称等が登録されているもの）をダウンロードしてフラッシュメモリ 2 2にも記憶させる。

15

また、C P U 1 9は、キーワードの自動抽出処理を行う前提として、ユーザーの選局操作時やユーザーの録画予約操作に基づく録画時に多重分離器 1 4から送られたE P G情報の packets をフラッシュメモリ 2 2に記憶させる。

20

図 3，図 4は、C P U 1 9が実行するキーワードの自動抽出処理を示すフローチャートである。このうち、図 3は、タイトル文字列情報からキーワードを抽出する処理であり、最初に、フラッシュメモリ 2 2に記憶させたE P G情報の中から、タイトル文字列情報を取り出す（ステップ S 1）。

25

続いて、そのタイトル文字列情報が示す複数の番組のタイトルから、タイトル用キーワード辞書に登録されている文字列（‘ゴルフ’，‘サッカー’，‘温泉’，‘囲碁’，‘将棋’，‘映画’といったサ

ブジャンルを示す文字列等)を探す。そして、それらの番組のタイトルのうち、このタイトル用キーワード辞書に登録されている文字列が含まれているタイトルの文字列全体を、キーワードの抽出対象とする(ステップS2)。

- 5 続いて、ステップS2でキーワードの抽出対象としたタイトルのうち、タイトル用除外文字列辞書に登録されている文字列(‘映画’、‘BS’等)の部分をスペースで置換する(ステップS3)。

続いて、ステップS3を経たタイトルの文字列から、図5に示すようなタイトル用の抽出ルールでキーワードを抽出する(ステ

- 10 ップS4)。

このタイトル用抽出ルールでは、そのタイトルの文字列が平仮名、カタカナ、漢字、数字、アルファベット以外の特殊文字(スペースや×や「」等)で区切られていない場合は、そのタイトルの文字列をそのままキーワードとして抽出する。他方、そのタイトルの文字列がこうした特殊文字で区切られている場合は、特殊

- 15 文字で区切られている各文字列のうちの2文字以上の文字列をそれぞれキーワードとして抽出する。
- 但し、‘・’(中点)は特殊文字としては扱わない。そして、キーワードとして抽出した文字列の先頭か末尾に‘・’(中点)が存在する場合には、‘・’(中点)を除いた部分をキーワードとする。

- 20 最後に、ステップS4で抽出したキーワードを、タイトル文字列情報中のキーワードのリストとしてフラッシュメモリ22に記憶させる(ステップS5)。

- 次に、図4は、詳細文字列情報からキーワードを抽出する処理
- 25 であり、最初に、フラッシュメモリ22に記憶させたEPG情報の中から、詳細文字列情報を取り出す(ステップS11)。

続いて、その詳細文字列情報から、詳細情報用キーワード辞書に登録されている文字列(有名な人の名称等)を探す。そして、

その詳細文字列情報のうち、この詳細情報用キーワード辞書に登録されている文字列をキーワードとして抽出するとともに、その文字列の部分を半角スペースで置換する（ステップ S 1 2）。

- 5 続いて、ステップ S 1 2 を経た詳細文字列情報の文字列のうち、
詳細情報用除外文字列辞書に登録されている文字列（‘ゲスト’、
‘以上’、‘監督’等）の部分を半角スペースで置換する（ステップ S 1 3）。

- 10 続いて、ステップ S 1 3 を経た詳細文字列情報の文字列から、
図 6 に示すような詳細文字列情報用の抽出ルールでキーワードを
抽出する（ステップ S 1 4）。

この詳細文字列情報用抽出ルールでは、基本的には、平仮名とカタカナと漢字と数字とアルファベットとその他の字種の文字とを互いに分離する字種切り法を利用する。

- 15 但し、カタカナとアルファベットとは同一の字種として扱う（分離しない）。また、‘・’（中点）は、その直前の文字がカタカナ、アルファベットである場合にはそれぞれカタカナ、アルファベットとして扱う（分離しない）。

- 20 そして、分離した各文字列のうち、平仮名のみの文字列、2文字以下の漢字のみの文字列、6文字以上の漢字のみの文字列を除いた文字列を、それぞれキーワードとして抽出する。但し、キーワードとして抽出した文字列の先頭か末尾に‘・’（中点）が存在する場合には、‘・’（中点）を除いた部分をキーワードとする。

- 25 最後に、ステップ S 1 2 で抽出したキーワードと、ステップ S 1 4 で抽出したキーワードを、詳細文字列情報中のキーワードのリストとしてフラッシュメモリ 2 2 に記憶させる（ステップ S 1 5）。

次に、この番組記録再生装置 2 において番組検索のためのキーワードが抽出される様子を、具体例を挙げて説明する。

ユーザーの選局操作時やユーザーの録画予約操作に基づく録画時に多重分離器 14 から送られてフラッシュメモリ 22 に記憶された EPG 情報中のタイトル文字列情報に、例えば次のようなタイトルが含まれていたとする（但し□□、△△はプロ野球チーム名である）。

愛のから騒ぎ

プロ野球中継 □□×△△

BS 映画「スペース・ウォーズ」

すると、図 3 の処理では、‘愛’、‘プロ野球’、‘映画’という文字列がタイトル用キーワード辞書に登録されているので、ステップ S2 で、これらの各タイトルについて、それぞれそのタイトルの文字列全体がキーワードの抽出対象となる。

そして、これらのタイトルのうち、BS 映画「スペース・ウォーズ」については、ステップ S3 で、‘BS’の部分と‘映画’の部分とがスペースで置換される。

また、これらのタイトルのうち、プロ野球中継 □□×△△については、‘プロ野球中継’と‘□□’との間にスペース（特殊記号）が存在し、‘□□’と‘△△’との間にも×（特殊記号）が存在するので、ステップ S4 で、文字列‘プロ野球中継’、‘□□’、‘△△’がそれぞれキーワードとして抽出される。

また、これらのタイトルのうち、‘BS’、‘映画’の部分をスペース置換された「スペース・ウォーズ」については、「」（特殊記号）で区切られており、また‘・’（中点）は特殊記号として扱わないので、ステップ S4 で、本来の映画のタイトルそのものである‘スペース・ウォーズ’がキーワードとして抽出される。

また、これらのタイトルのうち、愛のから騒ぎは、特殊記号で区切られていないので、ステップ S4 で、タイトルそのものである‘愛のから騒ぎ’がキーワードとして抽出される。

したがって、ステップ S 5 では、以下の文字列が番組検索用のキーワードとしてフラッシュメモリ 22 に記憶される（前述のように□□，△△はプロ野球チーム名である）。

愛のから騒ぎ

5 プロ野球中継

□□

△△

スペース・ウォーズ

10 このようにして、愛のから騒ぎ，スペース・ウォーズというように特殊文字で区切られていないタイトルについては、図 3 の処理により、そのタイトルに含まれる複数の文字列がばらばらのキーワードとして抽出されることなく、そのタイトルそのものがそのままの形でキーワードとして抽出される。

15 こうした特殊文字で区切られていないタイトルは、そのタイトルに含まれる‘愛’，‘スペース’等の個々の文字列は意味が広すぎて番組検索のためのキーワードとしてあまり役立たず（検索結果が非常に多くなり）、タイトルそのものとしてはじめて番組の効率的な検索のためのキーワードとして役立つことが多い。したがって、ユーザーは、抽出されたキーワード（タイトルそのもの）
20 を用いて、番組を効率的に検索することができるようになる。

また、スペース・ウォーズという映画のタイトル文字列については、タイトル文字列情報においてこのタイトルに付加されていた‘BS’，‘映画’というような番組検索のためには一般的過ぎる文字列がキーワードに含まれていないとともに、このタイトル
25 文字列情報においてこのタイトルを囲っていた「」もキーワードに含まれていない。したがって、ユーザーは番組を効率的に検索することができるようになる。

また一方では、プロ野球中継 □□×△△というように特殊文

字（スペースや‘×’）で区切られているタイトルについては、図 3 の処理により、特殊文字で区切られている個々の文字列であるプロ野球中継、□□、△△がそれぞれキーワードとして抽出される。

- 5 こうした特殊文字で区切られているタイトルは、その特殊文字で区切られている個々の文字列がそれぞれ番組検索のためのキーワードとして役立ち、タイトルそのものとしては限定されすぎて番組検索のためのキーワードとしてあまり役立たない（対戦チーム（□□や△△の具体名）が異なれば別のタイトルになってしまうので、検索結果がゼロまたは非常に少なくなる）ことが多い。したがって、ユーザーは、抽出されたキーワード（特殊文字で区切られている個々の文字列）を用いて、やはり番組を効率的に検索することができるようになる。
- 10

- 他方、図 4 の処理では、フラッシュメモリ 22 に記憶された EPG 情報中のこれらのタイトルの番組の詳細文字列情報から、詳細情報用キーワード辞書に登録されている有名人（愛のから騒ぎという番組の司会者、ゲストや、映画スペース・ウォーズに出演している俳優）の名称等が、ステップ S 12 でキーワードとして抽出される。
- 15

- 20 その際、苗字が漢字で名前が平仮名やカタカナの有名人の名称（例えば石田あかりという名称）もこの詳細情報用キーワード辞書に登録されているので、そうした有名人の名称もキーワードとして抽出される。

- また、インターネット経由でダウンロードした最新の詳細情報用キーワード辞書も用いるので、最近有名になったばかりの人の名称もキーワードとして抽出される。
- 25

また、その詳細文字列情報のうち、その有名人の名称等の部分と、詳細情報用除外文字列辞書に登録されている文字列（‘ゲスト’，

‘以上’，‘監督’等）の部分とが、ステップ S 1 2 及び S 1 3 で半角スペースに置換される。

そして、このスペース置換された詳細文字列情報の文字列から、ステップ S 1 4 で、図 6 に示したルールによってキーワードが抽出される。

その際、カタカナとアルファベットとは同一の字種として扱われるとともに‘・’（中点）はその直前の文字がカタカナ，アルファベットである場合にはそれぞれカタカナ，アルファベットとして扱われるので、名前と苗字との間に‘・’（中点）が挿入された外国人名（例えば B・ドゥーリー）もキーワードとして抽出される。

また、最新の詳細情報用キーワード辞書にもまだ登録されていない人（例えばデビューしたばかりの無名の芸能人）の名称でも、平仮名のみの名称や 2 文字以下の漢字のみの名称や 6 文字以上の漢字のみの名称（すなわち人名としてあまりなさそうな名称）でなければキーワードとして抽出される。

また、‘ゲスト’，‘以上’，‘監督’といったような番組検索のためには不適切な文字列は、スペース置換されているのでキーワードとして抽出されることはない。

これにより、ステップ S 1 5 では、苗字が漢字で名前が平仮名やカタカナの有名人名や、最近有名になったばかりの人の名称や、名前がアルファベットで表記され苗字がカタカナで表記された外国人名や、名前と苗字との間に‘・’（中点）が挿入された外国人名も、番組検索用のキーワードとしてフラッシュメモリ 2 2 に記憶される。したがって、ユーザーは、抽出されたキーワードを用いて、番組を効率的に検索することができるようになる。

なお、図 3，図 4 の処理によってフラッシュメモリ 2 2 に記憶させたキーワードをユーザーが番組検索のために用いる方法とし

ては、例えば、リモコン 28 による所定の操作に基づき、CPU 19 が、番組検索用画面（キーワードを一覧表示するとともにユーザーがその中の所望のキーワードを選択して検索を指示するための画面）の映像信号を作成して映像信号処理回路 17，映像出力端子 29 を経て表示装置 3 に送るといったような、適宜の方法をとればよい。

10 以上のように、この番組記録再生装置 2 では、EPG 情報中のタイトル文字列情報からのキーワードの抽出と詳細文字列情報からのキーワードの抽出とを、それぞれの情報に合せて互いに異なるキーワード辞書とルールとで行うことにより、小さなサイズのプログラムや辞書で精度よくキーワードを抽出することができるようになっている。

15 これにより、CPU 19 の処理能力やメモリ（ROM 20 やフラッシュメモリ 22 等）の容量がさほど小さくなくても、EPG 情報中のタイトル文字列情報及び詳細文字列情報から、ユーザーが番組を検索するためのキーワードを効率よく且つ精度よく自動的に抽出することができるようになっている。

20 なお、以上の例では、デジタルテレビジョン放送の番組を記録・再生する装置に本発明を適用している。しかし、これに限らず、アナログテレビジョン放送の番組を記録・再生する番組記録再生装置にも本発明を適用してよいことはもちろんである。

25 図 7 は、本発明を適用したアナログテレビジョン放送用の番組記録再生装置のハードウェア構成を示すブロック図である。アンテナ 31 で受信されて番組記録再生装置 41 に入力したアナログ放送信号中の映像・音声信号は、チューナ 42 で周波数帯を選択され、MP EG エンコーダ 43 で符号化される。

テレビ番組の視聴時には、この符号化された映像・音声データは、MP EG デコーダ 47 で復号されて、番組記録再生装置 41

から表示装置 6 1 に送られる。

他方、テレビ番組の記録時には、M P E G エンコーダ 4 3 で符号化された映像・音声データは、バス 4 4 を介して主記憶装置 4 5 に送られて、主記憶装置 4 5 に記録される。

- 5 そして、再生時には、主記憶装置 4 5 から読み出された映像・音声データが、バス 4 4 を介して M P E G デコーダ 4 7 に送られ、M P E G デコーダ 4 7 で復号されて、番組記録再生装置 4 1 から表示装置 6 1 に送られる。

- 10 また、チューナ 4 2 で周波数帯を選択されたアナログ放送信号から、E P G 取得モジュール 4 6 で E P G 情報が取得される。この E P G 情報も、バス 4 4 を介して主記憶装置 4 5 に送られて、主記憶装置 4 5 に記憶される。

- 15 また、インターネット 7 1 と接続するための通信インターフェース 4 8 , R O M 4 9 , 主記憶装置 5 0 , 補助記憶装置 5 1 , M P E G デコーダ 4 7 が、互いにバス 5 2 で結ばれている。

- 20 この番組記録再生装置 4 1 でも、前述のようなタイトル用キーワード辞書、タイトル用除外文字列辞書、詳細情報用キーワード辞書及び詳細情報用除外文字列辞書が R O M 4 9 に格納されている（詳細情報用キーワード辞書に関してはインターネット経由で
25 専用のサイトから最新のものをダウンロードして補助記憶装置 5 1 にも記憶させる）とともに、番組記録再生装置 4 1 全体を制御する C P U 5 3 が、図 3 , 図 4 に示したのと同じキーワードの自動抽出処理をこれらの辞書及び主記憶装置 4 5 内の E P G 情報を用いて行い、抽出したキーワードを補助記憶装置 5 1 に記憶させる。

この番組記録再生装置 4 1 でも、図 1 , 図 2 の番組記録再生装置 2 について説明したのと全く同様にして、E P G 情報中のタイトル文字列情報からのキーワードの抽出と詳細文字列情報からの

キーワードの抽出とを、それぞれの情報に合せて互いに異なるキーワード辞書とルールとで行うことにより、小さなサイズのプログラムや辞書で精度よくキーワードを抽出することができる。

- 5 これにより、CPU 53 の処理能力やメモリ（ROM 49 や補助記憶装置 51 等）の容量がさほど大きくなくても、EPG 情報中のタイトル文字列情報及び詳細文字列情報から、ユーザーが番組を検索するためのキーワードを効率よく且つ精度よく自動的に抽出することができる。

- 10 また、以上の例では、表示装置とは別体となった番組記録再生装置に本発明を適用している。しかし、これに限らず、この番組記録再生装置と表示装置とが一体となったテレビジョン受信機や、番組の記録再生機能を有しないテレビジョン受信機にも本発明を適用してよい。

- 15 また、以上の例では、EPG 情報中の番組のタイトル文字列情報、詳細文字列情報からのキーワードの検索のために本発明を適用している。しかし、これに限らず、テレビジョン番組以外のコンテンツ（例えばインターネット経由で配信されるコンテンツ）のタイトル文字列情報、詳細文字列情報からのキーワードの検索のためにも本発明を適用してよい。

- 20 また、本発明は、以上の例に限らず、本発明の要旨を逸脱することなく、その他様々の構成をとりうることはもちろんである。

- 25 以上のように、本発明によれば、CPU の処理能力やメモリの容量がさほど大きくない家電製品でも、EPG 情報のような番組のタイトル文字列情報及び詳細文字列情報から、ユーザーが番組を検索するためのキーワードを、効率よく且つ精度よく自動的に抽出することができるという効果が得られる。

請 求 の 範 囲

1. コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第1のキーワード辞書を用いてキーワードの抽出を行う第1の抽出手段と、

- 5 前記コンテンツの詳細文字列情報から、人名が登録されている第2のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第2の抽出手段とを備えたことを特徴とするキーワード自動抽出装置。

- 10 2. 前記第1の抽出手段は、第1のキーワード辞書に登録されている文字列を含むタイトル文字列のうち、所定の除外文字列辞書に登録されている文字列を除外した部分からキーワードを抽出することを特徴とする請求の範囲第1項記載のキーワード自動抽出装置。

- 15 3. 前記第1の抽出手段は、第1のキーワード辞書に登録されている文字列を含むタイトル文字列のうち、平仮名、カタカナ、漢字、数字、アルファベット以外の特殊文字で区切られている文字列をキーワードとして抽出することを特徴とする請求の範囲第1項記載のキーワード自動抽出装置。

- 20 4. 前記第2の抽出手段は、前記第2のキーワード辞書を用いてキーワードを抽出した前記詳細文字列情報の残りの部分のうち、所定の除外文字列辞書に登録されている文字列を除外した部分から、字種切り法を利用したキーワードの抽出を行うことを特徴とする請求の範囲第1項記載のキーワード自動抽出装置。

- 25 5. 前記第2の抽出手段は、字種切り法を利用しつつ、カタカナとアルファベットとを同一字種として扱うとともに、‘・’（中点）は、その直前の文字がカタカナ、アルファベットである場合にはそれぞれカタカナ、アルファベットとして扱うことを特徴とする請求の範囲第1項記載のキーワード自動抽出装置。

6. 前記第2のキーワード辞書をネットワーク経由でダウンロードする手段

をさらに備え、前記第2の抽出手段は前記ダウンロードされた第2のキーワード辞書を用いることを特徴とする請求の範囲第1項記載のキーワード自動抽出装置。

7. コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第1のキーワード辞書を用いてキーワードの抽出を行う第1のステップと、

10 前記コンテンツの詳細文字列情報から、人名が登録されている第2のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第2のステップとを有することを特徴とするキーワード自動抽出方法。

15 8. 前記第1のステップで、第1のキーワード辞書に登録されている文字列を含むタイトル文字列のうち、所定の除外文字列辞書に登録されている文字列を除外した部分からキーワードを抽出することを特徴とする請求の範囲第7項記載のキーワード自動抽出方法。

20 9. 前記第1のステップで、第1のキーワード辞書に登録されている文字列を含むタイトル文字列のうち、平仮名、カタカナ、漢字、数字、アルファベット以外の特殊文字で区切られている文字列をキーワードとして抽出することを特徴とする請求の範囲第7項記載のキーワード自動抽出方法。

25 10. 前記第2のステップで、前記第2のキーワード辞書を用いてキーワードを抽出した前記詳細文字列情報の残りの部分のうち、所定の除外文字列辞書に登録されている文字列を除外した部分から、字種切り法を利用したキーワードの抽出を行うことを特徴とする請求の範囲第7項記載のキーワード自動抽出方法。

11. 前記第2のステップで、字種切り法を利用しつつ、カタカ

ナとアルファベットとを同一字種として扱うとともに、‘・’（中点）は、その直前の文字がカタカナ，アルファベットである場合にはそれぞれカタカナ，アルファベットとして扱うことを特徴とする請求の範囲第 7 項記載のキーワード自動抽出方法。

- 5 1 2 . 前記第 2 のキーワード辞書をネットワーク経由でダウンロードするステップ

をさらに有し、前記第 2 のステップでは前記ダウンロードした第 2 のキーワード辞書を用いることを特徴とする請求の範囲第 7 項記載のキーワード自動抽出方法。

- 10 1 3 . キーワード自動抽出装置のプログラムであって、

コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第 1 のキーワード辞書を用いてキーワードの抽出を行う第 1 の抽出ステップと、

- 15 前記コンテンツの詳細文字列情報から、人名が登録されている第 2 のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第 2 の抽出ステップとを含むことを特徴とするコンピュータが読み取り可能なプログラムが記録されている記録媒体。

- 1 4 . キーワード自動抽出装置を制御するコンピュータに、

- 20 コンテンツのタイトル文字列情報から、サブジャンルを示す文字列が登録されている第 1 のキーワード辞書を用いてキーワードの抽出を行う第 1 の抽出ステップと、

- 25 前記コンテンツの詳細文字列情報から、人名が登録されている第 2 のキーワード辞書を用いたキーワードの抽出と、字種切り法を利用したキーワードの抽出とを行う第 2 の抽出ステップとを実行させることを特徴とするプログラム。

FIG. 1

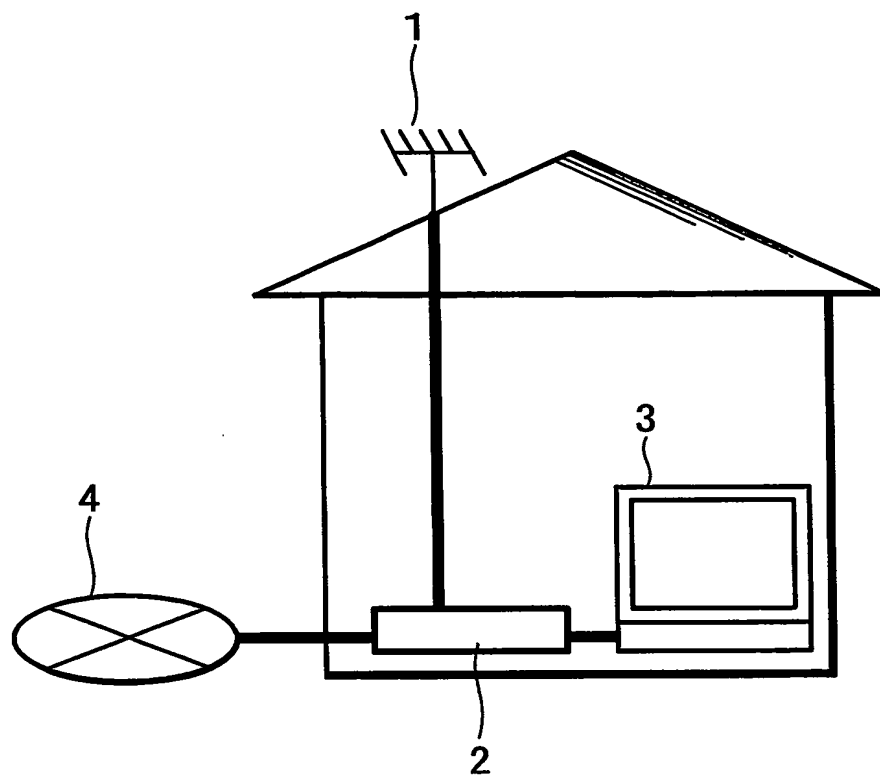


FIG. 2

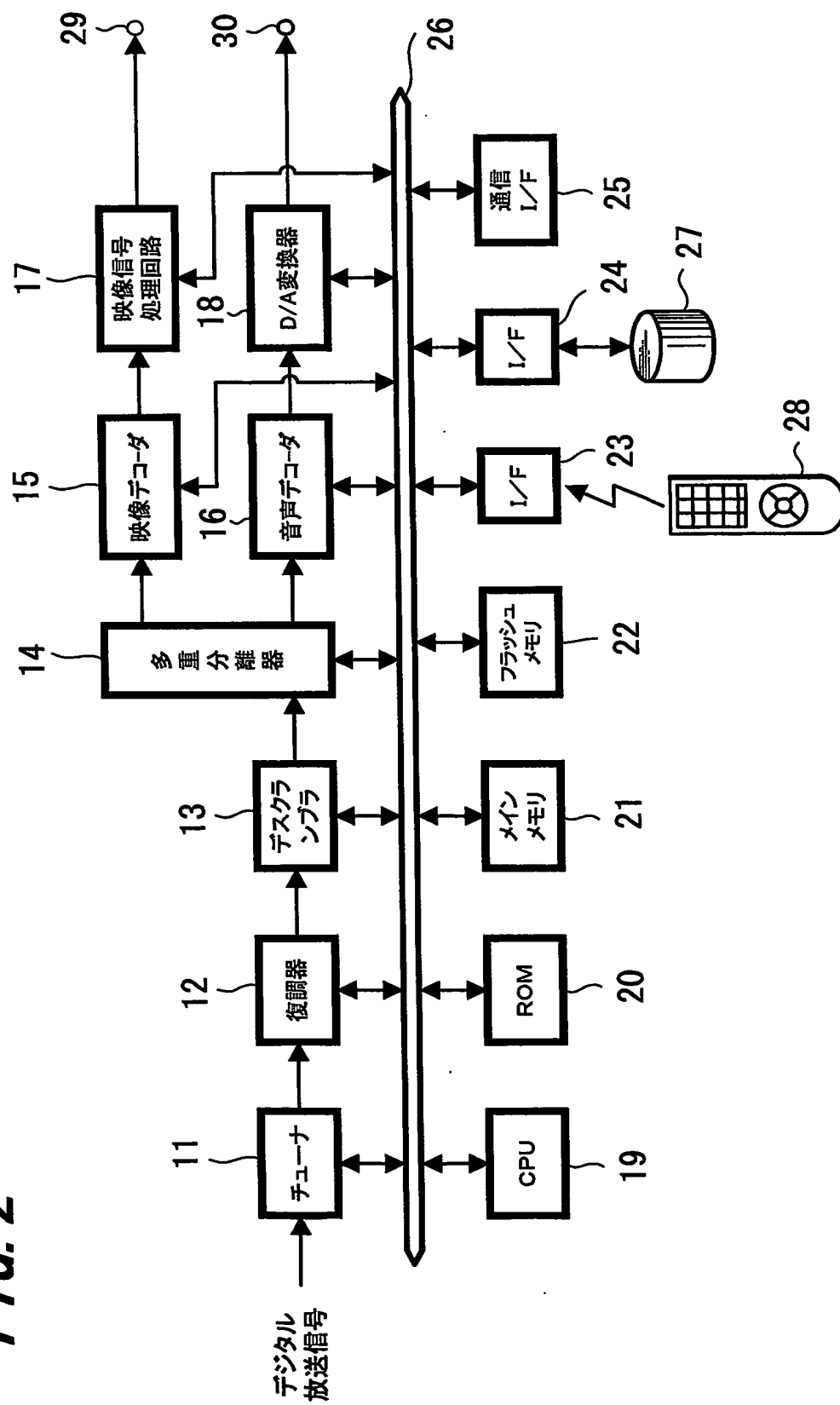


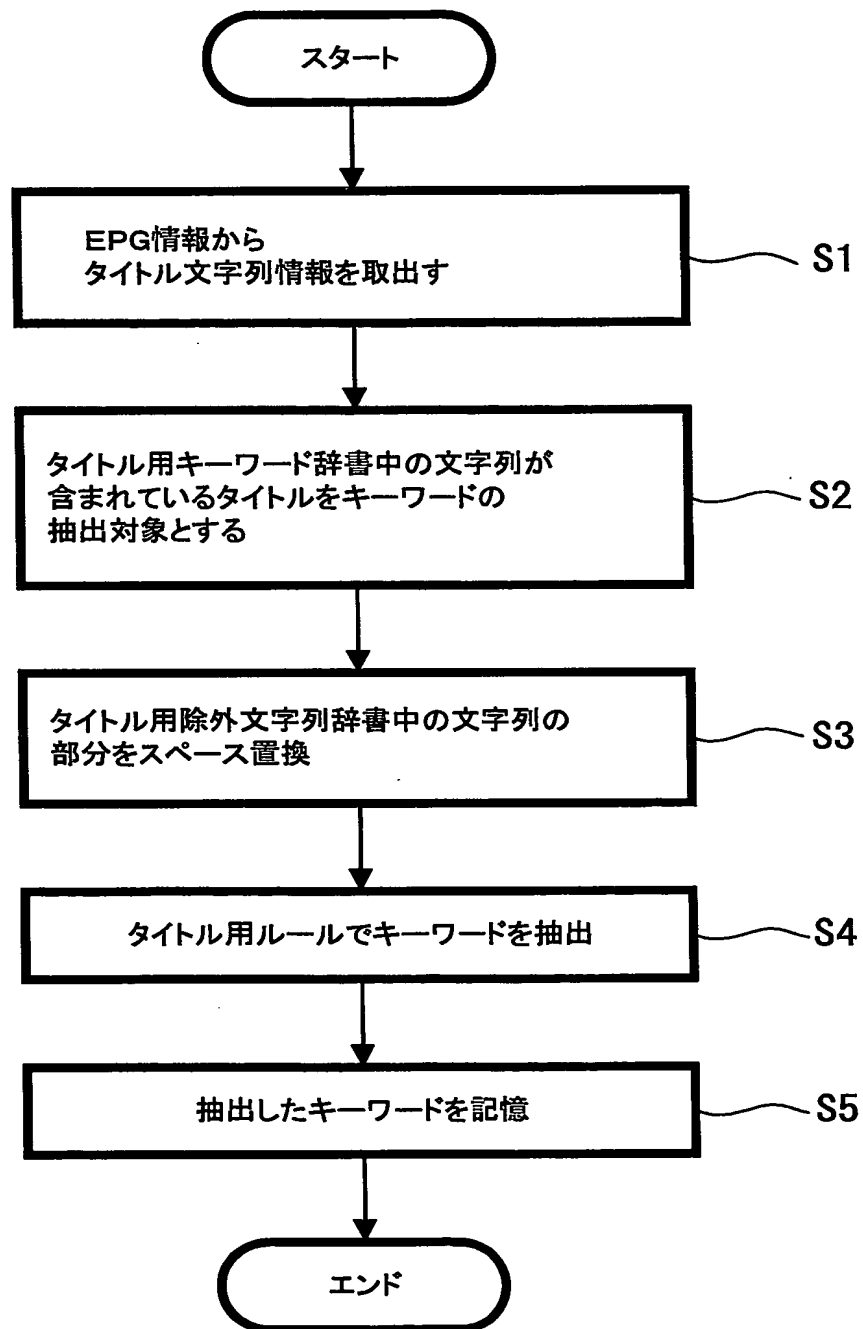
FIG. 3

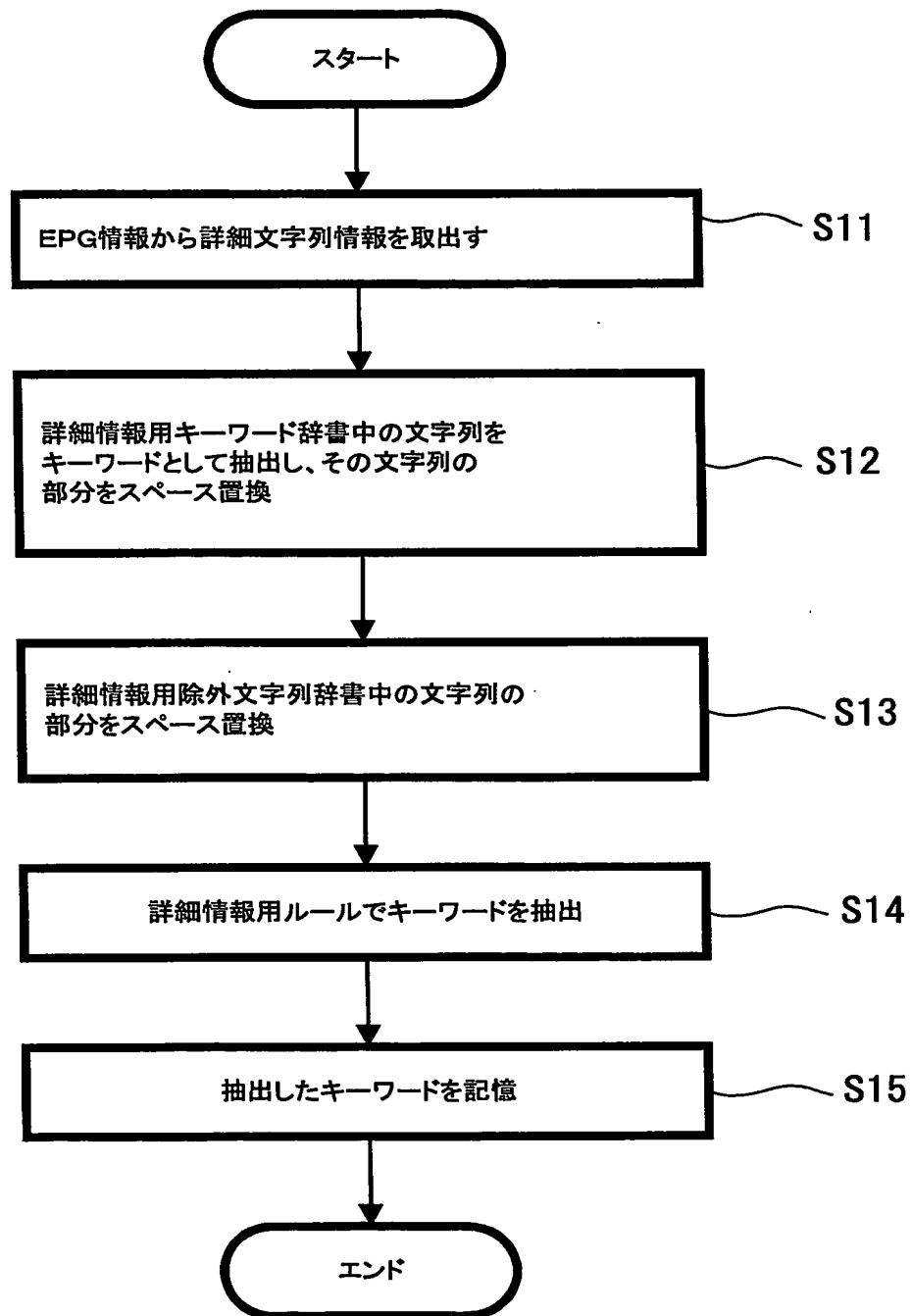
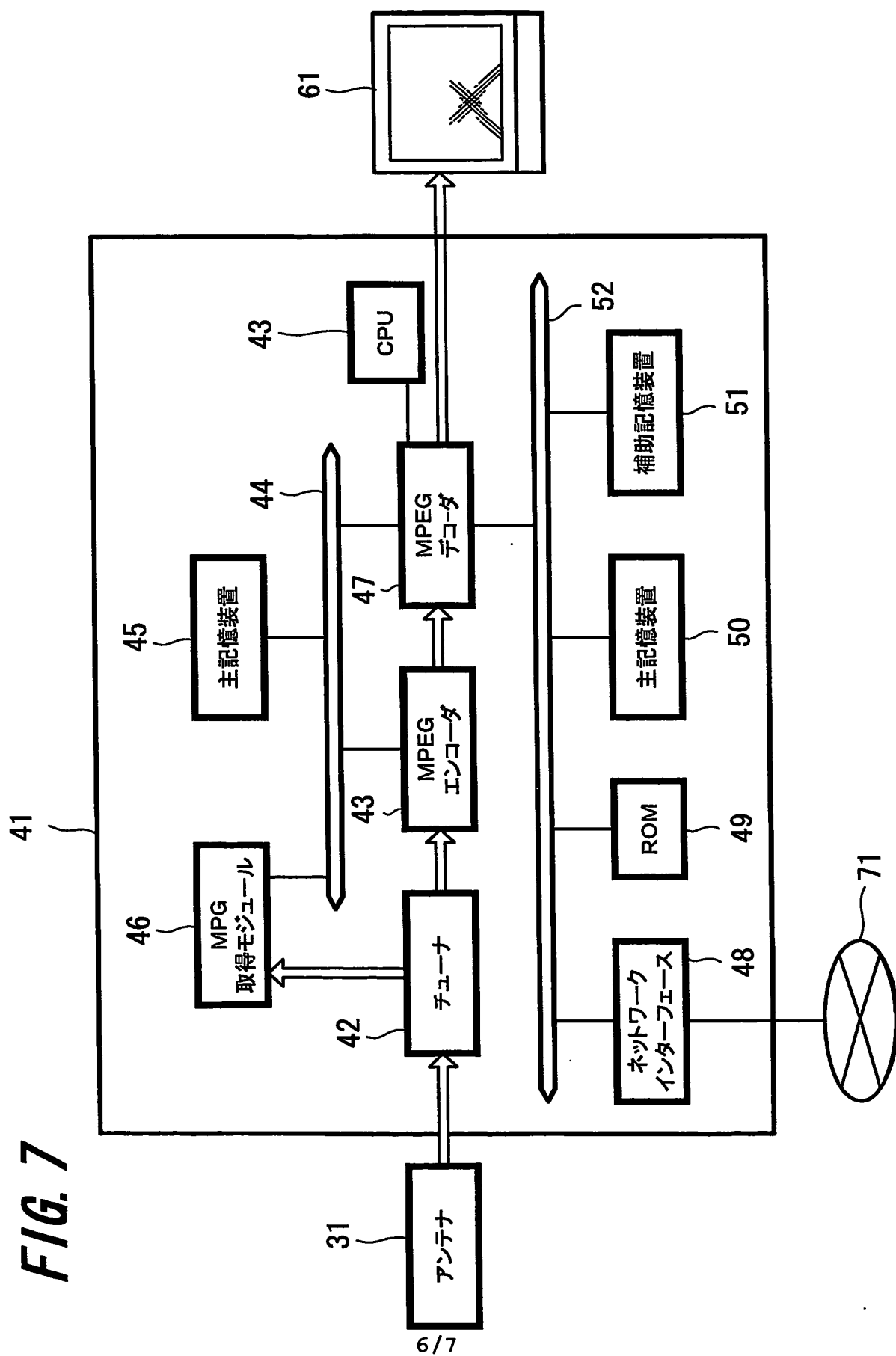
FIG. 4

FIG. 5

特殊文字で区切られていないタイトルの 文字列をそのままキーワードとして抽出
特殊文字で区切られているタイトルの 各文字列のうち、2文字以上の文字列を キーワードとして抽出
'・' (中点) は特殊文字として扱わない。 抽出した文字列の先頭か末尾に、'・' が 存在する場合には、'・' を除いた部分を キーワードとする

FIG. 6

字種切り法を利用
カタカナとアルファベットは同一字種として扱う
'・' は、直前の文字がカタカナ、アルファベット のとき、それぞれカタカナ、アルファベットとして 扱う
分離した文字列のうち、平仮名のみ、2文字以下の 漢字の み、6文字以上の漢字のみを除いた文字列を キーワードとして抽出



引用符号の説明

- 1, 31…アンテナ
- 2, 41…番組記録再生装置
- 3, 61…表示装置
- 11, 42…チューナ
- 12…復調器
- 13…デスクランブラ
- 14…多重分離器
- 15…映像デコーダ
- 16…音声デコーダ
- 17…映像信号処理回路
- 18…D/A変換器
- 19…CPU
- 20, 49…ROM
- 21…RAM
- 22…フラッシュメモリ
- 23…リモートコントローラ用のインターフェース
- 24…HDD用のインターフェース
- 25, 48…インターネット接続用の通信インターフェース
- 26…システムバス
- 27…HDD
- 28…リモコン
- 43…MPEGエンコーダ
- 44, 52…バス
- 45, 50…主記憶装置
- 46…EPG取得モジュール
- 47…MPEGデコーダ
- 50…主記憶装置
- 51…補助記憶装置

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/09678

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Toroku Jitsuyo Shinan Koho	1994-2003
Kokai Jitsuyo Shinan Koho	1971-2003	Jitsuyo Shinan Toroku Koho	1996-2003

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS), WPI, INSPEC (DIALOG)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Yuichiro AOKI et al., "Joho Kensaku System DATA-710", NEC Giho, 31 October, 1998 (31.10.98), Vol.41, No.12, pages 33 to 39; in particular, page 35, left column to page 36, left column, Figs. 4 to 6	1-14
Y	JP 2001-75959 A (Matsushita Electric Industrial Co., Ltd.), 23 March, 2001 (23.03.01), Column 5, line 8 to column 6, line 32 (Family: none)	1-14
Y	JP 8-10452 B2 (Nippon Telegraph And Telephone Corp.), 31 January, 1996 (31.01.96), Full text; all drawings (Family: none)	1-14

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
28 August, 2003 (28.08.03)Date of mailing of the international search report
07 October, 2003 (07.10.03)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/09678

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Ken'ichi HINATSU, "JICST ni okeru Keyword Jido Chushutsu System no Shiyo", The Journal of Information Science and Technology Association, 01 November, 1992 (01.11.92), Vol.42, No.11, pages 1051 to 1057	5,11
Y	JP 10-198667 A (Casio Computer Co., Ltd.), 31 July, 1998 (31.07.98), Full text; all drawings (Family: none)	6,12

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2003年
 日本国登録実用新案公報 1994-2003年
 日本国実用新案登録公報 1996-2003年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS), WPI, INSPEC (DIALOG)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	青木祐一郎 他, 情報検索システムDATA-710, NEC技報, 1998. 10. 31, 第41巻, 第12号, p. 33-39, とくに第35頁左欄~第36頁左欄, 図4~6	1-14
Y	JP 2001-75959 A (松下電器産業株式会社) 2001. 03. 23, 第5欄第8行~第6欄第32行 (ファミリーなし)	1-14
Y	JP 8-10452 B2 (日本電信電話株式会社) 1996. 01. 31, 全文, 全図 (ファミリーなし)	1-14

☒ C欄の続きにも文献が列举されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」 同一パテントファミリー文献

国際調査を完了した日

28. 08. 03

国際調査報告の発送日

07.10.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

深津 始

5M

9383

電話番号 03-3581-1101 内線 3597

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	日夏健一, J I C S Tにおけるキーワード自動抽出システムの使 用, 情報の科学と技術, 1992. 11. 01, 第42巻, 第11 号, p. 1051-1057	5, 11
Y	J P 10-198667 A (カシオ計算機株式会社) 199 8. 07. 31, 全文, 全図 (ファミリーなし)	6, 12